

# An Assessment of Family History Information Captured in an Electronic Health Record

Fernanda Polubriaginof, MD<sup>1</sup>, Nicholas P. Tatonetti, PhD,<sup>1</sup> David K. Vawdrey, PhD<sup>2,1</sup>

<sup>1</sup>Department of Biomedical Informatics, Columbia University, New York, NY, USA;

<sup>2</sup>New York-Presbyterian Hospital, New York, NY, USA

## Abstract

*Family history is considered a core element of clinical care. In this study we assessed the quality of family history data captured in an established commercial electronic health record (EHR) at a large academic medical center. Because the EHR had no centralized location to store family history information, it was collected as part of clinical notes in structured or free-text format. We analyzed differences between 10,000 free-text and 9,121 structured family history observations. Each observation was classified according to disease presence/absence and family member affected (e.g., father, mother, etc.). The structured notes did not collect a complete family history as defined by standards endorsed by the U.S. Agency for Healthcare Research and Quality; the free-text notes contained more information than the structured notes, but still not enough to be considered “complete.” Several barriers remain for collecting complete, useful family history data in electronic health records.*

## Introduction

Family history has always been considered “a core element of clinical care.”<sup>1</sup> It plays an important role in personalized medicine, being a free genetic tool that almost every patient has access to.<sup>2</sup> Since the Human Genome Project, new genomic tools have been described,<sup>3</sup> but family history is critical for identifying patients that may benefit from more intensive screening. Family history provides information that enables individualized disease diagnosis, treatment and prevention.

Several studies have shown that family history can help predict the risk of a patient developing certain diseases such as breast cancer, colorectal cancer, ovarian cancer, osteoporosis, cardiovascular disease, psychiatric disorders, and diabetes. Knowing that a patient is at increased risk of developing a disease based on family history enables disease prevention that can vary from intensive screening to prophylactic surgery, early diagnosis and/or early and tailored treatment. For example, current guidelines from the American Cancer Society define criteria for MRI eligibility in addition to mammography for breast cancer screening<sup>4,5</sup>. The guidelines recommend that patients that have lifetime risk of breast cancer greater than 20% by BRCAPRO<sup>6</sup>, Tyrer-Cuzick<sup>7</sup> or Claus<sup>8</sup> models should have screening MRI in addition to mammography. Each of these models was developed using different methods, different populations and different risk factors and each of them was developed to predict different outcomes, but all of them heavily rely on family history and presence of risk factors<sup>9</sup>. This is just one of many examples of the importance of family history information for clinical care.

Since several guidelines and models exist to estimate a patient’s risk for different conditions, clinical decision support systems have been developed to facilitate guided personalized medicine. One of the earliest research studies about the use of clinical decision support (CDS) systems in personalized medicine used was conducted by Emery and colleagues in 1999. The study identified that the CDS systems available were not appropriate for use in a primary care setting.<sup>10</sup> To address this problem, they developed a system to record and interpret family history data in the primary care clinic. The system included family history relevant to breast, ovarian and colorectal cancer.<sup>11</sup> Over time, other clinical decision support systems were developed to manage other types of cancers such as colorectal cancer and Lynch syndrome. All these systems used family history information to provide risk assessment.<sup>12</sup>

Currently, the U.S. Preventive Services Task Force (USPSTF) includes risk assessment based on family history for some conditions, demonstrating the importance of family history in clinical care. Conditions that include family history information as part of the USPSTF are screening for BRCA mutation and BRCA-related cancers<sup>13</sup>, osteoporosis<sup>14</sup> and lipid disorders in adults.<sup>15,16</sup> A 2007 report commissioned by the Agency for Healthcare Research and Quality (AHRQ) recommended that collection of family history information should include diseases in first-degree relatives and second-degree relatives from both the maternal and paternal side, the relative’s age at the time of disease diagnosis, and each relative’s race and ethnicity.<sup>25</sup>

In electronic health records (EHRs), family history information may be collected in a structured or narrative (i.e., “free-text”) format. On one hand, structured documentation is ideal for data reuse since it may be coded and standardized. On the other hand, when clinicians are providing care, flexibility, efficiency, quality, expressivity are important.<sup>17,18</sup> The purpose of this study was to understand how family history data were captured in an established commercial EHR system at a large academic medical center, and to assess the quality of the data collected.

## Methods

With Institutional Review Board approval, we conducted a retrospective analysis of data from the Allscripts EHR (Allscripts Corp., Chicago IL) used at NewYork-Presbyterian Hospital/Columbia University Medical Center since 2004.

Our study focused on the difference between family history data collected using structured fields vs. free-text. Each note template in the EHR contained one or more “observation” data elements. An observation could be a text box, a Boolean (e.g., a checkbox or radio button), or numeric value. Text boxes could be fully free-text, or they could be constrained to allow only items from a predefined list, such as the words “low,” “medium,” or “high.” Observations had an internal code and description specified using a configuration tool in the EHR. While our EHR vendor provided some predefined observations, the vast majority were locally defined and did not comport with any existing standard terminology.

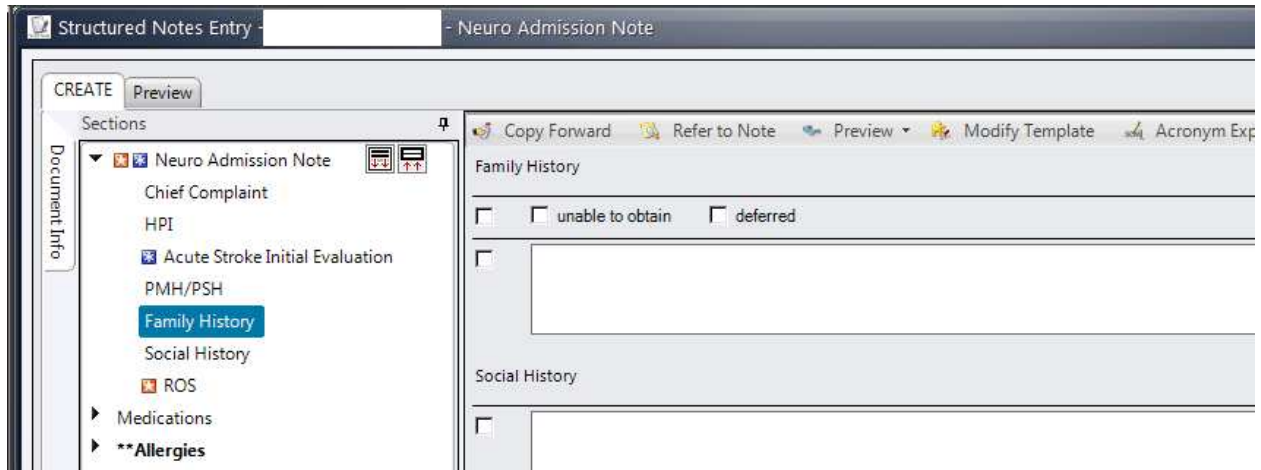
The EHR system contained 1,560 active templates for documentation. Each of these templates contained one to several hundred discrete observations. There were 140,038 observations defined in the EHR; of those, 653 had an internal code containing the words “fam hx” OR “family hist.” We identified the note templates that contained these observations and queried the EHR database to identify the number of times each note template was used, as well as the number of unique patients who had at least one of these observations recorded. The number of times each note template was used varied from 1 to 79,505, and number of unique patients varied from 1 to 67,276 for each note template. The note templates that contained the most commonly used free-text and structured text observations were selected for analysis.

The most-used note template that contained structured family history observations was the Ambulatory OB/GYN Antepartum Record (Figure 1). This note template was used in our institution for obstetric patients in the institution’s ambulatory care network. Overall, this note template was used 79,505 times for 67,276 unique patients. The most-used free-text family history observation was the Neurology Admission Note (Figure 2). This note was used for every patient admitted to the neurology service. The Neurology Admission Note was used 49,656 times for 22,642 unique patients.

The screenshot shows the 'Structured Notes Entry' interface for an 'Amb OB/GYN Antepartum Record'. On the left is a 'Document Info' sidebar with a tree view of sections including Informant/Interpreter, Patient Sex/Age, Subjective, Allergies, Health Issues, Care Providers, Present Pregnancy, Past Pregnancies, and Past Medical History (with sub-sections for Gynecologic, Medical History, Surgery, Alcohol/Drug/Tobacco Usage, Family History, and Screening for Genetic Disorders). The main area contains a 'Family History' section with a table of radio button options for 'Yes', 'No', and 'N/A' for each condition, followed by a 'Remarks/Comments' text box.

Condition	Yes	No	N/A	Text Box
Multiple Preg:	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="text"/>
Diabetes:	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="text"/>
Hypertension:	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="text"/>
Heart Disease:	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="text"/>
Renal Disease:	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="text"/>
Neuro/Psych Disease:	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="text"/>
Cancer: Breast/Gyn:	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="text"/>
Bleeding Disorder:	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="text"/>
Cancer: Other:	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="text"/>

**Figure 1.** Ambulatory OB/GYN Antepartum Record: the most-used template note that contained structured family history observations.



**Figure 2.** Neurology Admission Note: the most-used template note that contained free-text family history observation.

For both the Ambulatory OB/GYN Antepartum Record and the Neurology Admission Note templates, 10,000 family history observation entries were randomly selected from notes between 2007 and 2014. Manual annotation by a clinical expert (FP) was performed in all 10,000 free-text observations, as well as in all structured observations that occurred more than once (9,121 observations). Categories were identified based on the content of information in the observations and the standards endorsed by AHRQ<sup>25</sup> and were used to annotate both datasets. In total, 19,121 observations were manually annotated and assigned to a category. The categories used were: 1) presence of disease in specified relative(s), 2) presence of disease in unspecified relative(s), 3) absence of disease and 4) other (Table 1). The annotation results were compared between the datasets.

**Table 1.** Family history categories with definitions and examples.

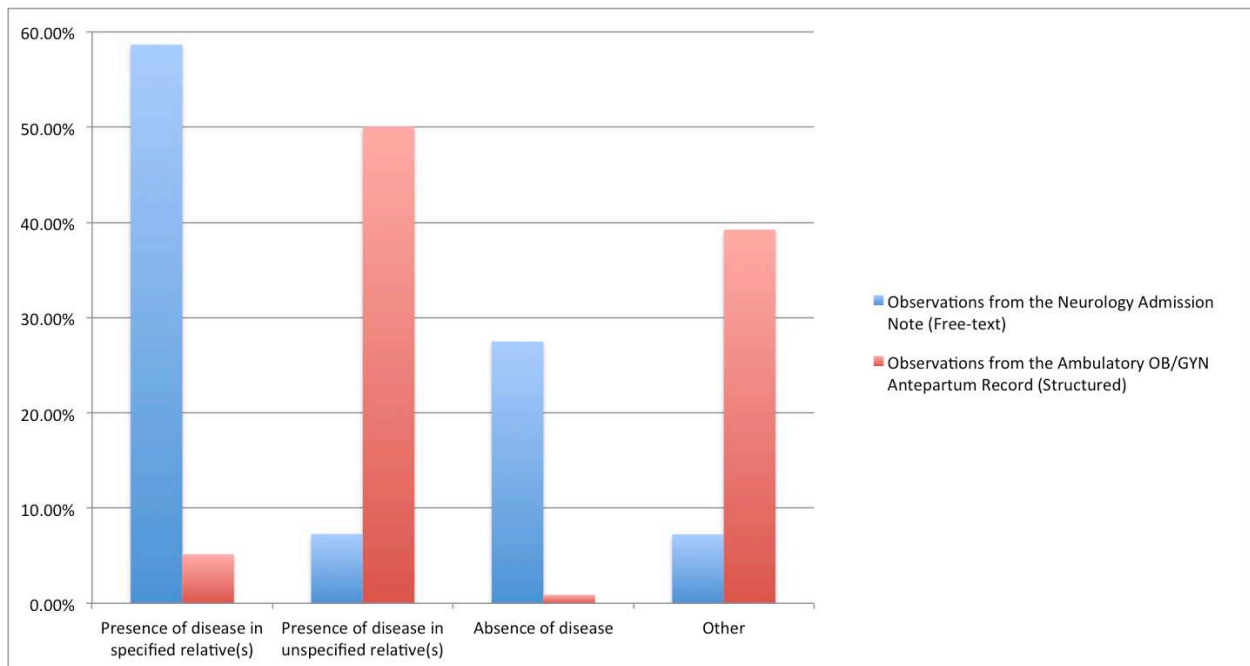
Family History Categories	Definition	Examples
<b>Presence of disease in specified relative(s)</b>	When family history of disease is reported paired with a relative	“Mother: hepatic cancer; Brother: colon cancer” “Hypertension Mother”
<b>Presence of disease in unspecified relative(s)</b>	When a family history of disease is reported by itself without affected relative information	“History of diabetes, hypertension, MI, strokes.” “Diabetes”
<b>Absence of disease</b>	When family history of disease is negated	“No dementia; No strokes.” “Diabetes Denies”
<b>Other</b>	Miscellaneous responses	“Non-contributory” “None” “no Arabic translator” “No family at bedside and pt nonverbal” “Pt adopted, unknown family history”

## Results

Figure 3 shows the results of the manual annotation of family history observations. In the category “Presence of disease in specified relative(s)”, 58.7% of the observations from the Neurology Admission Note (free-text) contained family history information including the disease as well as the relative affected. However, when analyzing data from the observations from the Ambulatory OB/GYN Antepartum Record (structured), only 5.2% contained information specifying the patient’s relative. In contrast, 7.3% of the observations from the Neurology Admission Note (free-text) contained information categorized as “Presence of disease in unspecified relative(s),” and 50.1% of the observations from the Ambulatory OB/GYN Antepartum Record (structured) captured this type of information (Figure 3).

Furthermore, 27.5% of the observations from the Neurology Admission Note (free-text) captured information about the absence of family history of a certain disease, while only 0.9% of the observations from the Ambulatory OB/GYN Antepartum Record (structured) captured information with this level of detail (Figure 3).

A large proportion (39.2%) of the observations from the Ambulatory OB/GYN Antepartum Record (structured) were classified as “Other.” The vast majority of these cases referred to family history described as “N/A.” Such description provides no information of the patient’s family history (Figure 3).



**Figure 3.** Comparison of categories from free-text and structured family history observations.

## Discussion

When analyzing family history data collected using structured vs. free-text data elements, the clinical expert annotations revealed that there was a considerable difference between the content of the family history information collected. The structured template was shown to be a poor way to collect family history since the relative affected by the condition was not captured. Other important details such as age of onset and vital status were also not being captured. In the template shown in Figure 1, text could be optionally entered in the free-text box but this field was stored separately in the database from the structured component. In Figure 1, it was also unclear what was meant by “N/A.” Possible interpretations were that this could indicate that a patient would not or could not inform, had no knowledge or even that such questions were not asked. Another problem was that rows could easily be left empty and it is unclear what that means. It could mean that the clinician did not ask about the family history of a certain condition, or that the patient did not know. Furthermore, the free-text box was used to capture relatives, deny presence conditions, or even to record other types of information such as age, type of cancer, etc. Notes that used the

free-text template were more comprehensive and often contained much useful information. However, not all of them contained pertinent and useful family history information.

Unlike some EHR implementations where family history information is kept in a centralized fashion, our institution did not have a centralized location to store this type of data. Family history data was collected and stored as part of the clinical notes. The fact that different healthcare providers collected this type of information in several different notes may have resulted in inconsistencies across notes. Our study discovered that our institution's EHR 653 observations related to family history. Many different note-writing templates were available and many included family history sections, but there was no standardized method for collecting this type of data. One benefit of having so many different ways of collecting family history data was that we could assess the quality of data based on characteristics such as structured items vs. free-text fields.

We also observed that free-text boxes in both notes were often used to capture information that was unrelated to family history. For example, indicators were included such as: "intubated," and "no family at bedside and pt nonverbal". These are important pieces of information about the patient but should not be reported as part of the family history section.

Despite the well-known and well-described importance of family history, several barriers exist in its collection and analysis, as well as in its use for personalized management based on patients' risk assessment. Barriers to collect family history can be classified in two major categories: clinician-related and patient-related.

Clinician-related barriers include lack of time to obtain, organize and analyze family history information; lack of resources and lack of reimbursement for such activity; underestimation of the value of family history data by the clinician; lack of expertise in obtaining and analyzing family history information; lack of standards for family history collection; and lack of clear guidelines to assess patient risk based on family history. The first, and perhaps the most critical barrier for family history collection is lack of time to obtain, organize and analyze family history information.<sup>2,19-23</sup> Obtaining complete and accurate family history information, organizing it in a pedigree and analyzing family history data is extremely time-consuming. Furthermore, it is not sufficient to collect family history from patients only once. It is important to regularly update family history information, analyze it, and reconcile conflicting information. A 1989 study surveying four genetic clinics reported that the time patients spent in the first consultation varied from 3–5.5 hours, with over half of this time spent before or after the patient's appointment.<sup>24</sup> A 2011 study demonstrated that while the majority of clinicians (77.5%) reported collecting cancer family history on their patients, only 26.0% included minimum adequate cancer family history.<sup>23</sup> Furthermore, 57.4% of clinicians updated family history information just once a year, and 22.2% of clinicians never updated family history information for their patients at all. When questioned about the barriers to collecting cancer family history, clinicians reported lack of time as the primary issue.<sup>23</sup> The study focused on cancer family history, but it demonstrated how challenging family history is to collect and maintain, in general.

Lack of resources and reimbursement for family history collection is another important barrier.<sup>19-22</sup> Clinicians are not reimbursed for the time spent on family history collection and risk assessment. In fact, in 2009, lack of incentives from the government was being described as one of the challenges prohibiting adequate collection of family history.<sup>20</sup> In addition to misaligned incentives, lack of standards has also been reported as a challenge in this area.<sup>20,21,26</sup> The lack of standards for data elements, terminology, structure, interoperability, and clinical decision support rules for family history data is a huge obstacle to implement it in the clinical workflow. This point is underscored by the existence of multiple EHR templates available to assist physicians in capturing family history data. Furthermore, limited knowledge and lack of expertise in obtaining and analyzing family history by clinicians is another barrier that has been described in several studies.<sup>2,20,21,23,26,27</sup>

There are also barriers to collecting accurate family history data on the patient side. These include uncertainty about biological family composition; uncertainty about the health history of family members; inaccuracies in patient recall, language-related and cultural factors. Clinicians often cite uncertainty about biological family composition as a challenge when collecting family histories, especially in cases where the patient is part of a large biological family.<sup>21,23,28</sup> Language-related and cultural factors can also be a factor that negatively affects collection.<sup>23</sup>

There are now several initiatives to facilitate and encourage the use of family history data. These initiatives are focused on the use of these data for precision medicine, where the need for accurate and detailed family history data is great. Three such initiatives are: Stage 2 of the federal "Meaningful Use" EHR financial incentive program, the U.S. Surgeon General's "My Family Health Portrait," and the HL7 Clinical Genomics Family History/Pedigree Model.

Stage 2 of the Meaningful Use program included a requirement of clinician's to use structured data entry for family history. Eligible hospitals had to have for 20% of their patients at least one structured family history data element, for at least one first-degree relative in the electronic health record.<sup>29</sup> As discussed above, lack of incentives to collect family history is an important barrier. The Meaningful Use program will be a strong incentive for U.S. hospitals to collect family history information. Although the determined measure of at least one structure data entry for at least one first-degree relative is far from what is considered complete family history, it is a start.

Secondly, since family history data collection is extremely time-consuming, innovative tools have been created to facilitate this process. Some are leveraging patient input of family history data. The U.S. Surgeon General's My Family Health Portrait<sup>30</sup>, a federal initiative to collect family history, is a website that allows patients to collect family history information and share their information with family members and healthcare providers. A study conducted in 2011 described that the average time taken to input family history information was 15 minutes, in a range from 3 to 45 minutes.<sup>31</sup> Instead of having a healthcare provider questioning patients about their family history, patients can enter their own data, saving clinician time--the major barrier for family history data collection. This practice also engage patients in their care and gives them time to review their family information and contact relatives and question them about information that they may not know. Engaging patients in this fashion encourages more accurate family history information. Of note, one advantage of using electronic questionnaires is that certain questions can be made mandatory, and branching logic can be employed. In contrast, in a clinical encounter, the doctor may forget to ask certain questions or may skip questions due to lack of time.

It is important to emphasize that to fully represent family history information, data representation must be multidimensional since it is necessary to not only capture the disease but also the relative affected, age of onset, and cancer type if applicable. Moreover, development of standards to support interoperability is essential for sharing data for clinical care and clinical research purposes. In the domain of family history, HL7 has a workgroup that specifically works on development of models for representing family history. The workgroup has developed the HL7 Clinical Genomics Family History (Pedigree) Model.<sup>32</sup> It is a standard for capturing data within a system as well as to transmit family history data between systems. It includes patient's family and familial relationships, diseases, genetic data and risk analysis. This HL7 standard is already used by the U.S. Surgeon General's My Family Health Portrait application, and we believe it will be important for EHR vendors and other stakeholders to adopt this standard

## **Limitations**

The study analyzed only observations contained in two note templates (out of a total of 1,560 available) in our EHR. One note template was used in the ambulatory care setting and the other template was used for hospital admissions. Although the templates were selected based on the fact that they were the most frequently used templates at our institution, it is unclear if analysis of other EHR templates would yield different conclusions.

## **Conclusion**

In summary, family history information is a valuable tool for personalized medicine. Computer-based tools can improve data collection, risk assessment and decision-making, facilitating effective use of family history in the clinic, engaging patients in their care, and therefore providing better personalized care. Our electronic health record vendors have not developed a user-friendly system to collect this type of data. Patient-facing tools for collecting family history data are not yet being used on a large scale. Numerous efforts have been made to collect family history data in the electronic format and to facilitate its use in the clinical setting, but several barriers remain unsolved. We have demonstrated that free-text observations were more comprehensive than structured observations; however neither was ideal for capturing patients' complete family history.

## **References**

1. Berg AO, Baird MA, Botkin JR, et al. National Institutes of Health State-of-the-Science Conference Statement: Family History and Improving Health. *Ann Intern Med.* 2009;151(12):872-877.
2. Guttmacher AE, Collins FS, Carmona RH. The family history--more important than ever. *N Engl J Med.* 2004;351(22):2333-2336.

3. Guttmacher AE, Collins FS. Welcome to the genomic era. *N Engl J Med*. 2003;349(10):996-998.
4. Saslow D, Boetes C, Burke W, et al. American Cancer Society guidelines for breast screening with MRI as an adjunct to mammography. *CA Cancer J Clin*. 2007;57(2):75-89.
5. Smith RA, Cokkinides V, Brawley OW. Cancer screening in the United States, 2012: A review of current American Cancer Society guidelines and current issues in cancer screening. *CA Cancer J Clin*. 2012.
6. Berry DA, Parmigiani G, Sanchez J, Schildkraut J, Winer E. Probability of carrying a mutation of breast-ovarian cancer gene BRCA1 based on family history. *J Natl Cancer Inst*. 1997;89(3):227-238.
7. Tyrer J, Duffy SW, Cuzick J. A breast cancer prediction model incorporating familial and personal risk factors. *Stat Med*. 2004;23(7):1111-1130.
8. Claus EB, Risch N, Thompson WD. Autosomal dominant inheritance of early-onset breast cancer. Implications for risk prediction. *Cancer*. 1994;73(3):643-651.
9. Ozanne EM, Drohan B, Bosinoff P, et al. Which risk model to use? Clinical implications of the ACS MRI screening guidelines. *Cancer Epidemiol Biomarkers Prev*. 2013;22(1):146-149.
10. Emery J. Computer support for genetic advice in primary care. *Br J Gen Pract*. 1999;49(444):572-575.
11. Emery J, Walton R, Coulson A, Glasspool D, Ziebland S, Fox J. Computer support for recording and interpreting family histories of breast and ovarian cancer in primary care (RAGs): qualitative evaluation with simulated patients. *BMJ*. 1999;319(7201):32-36.
12. Welch BM, Kawamoto K. Clinical decision support for genetically guided personalized medicine: a systematic review. *J Am Med Inform Assoc*. 2013;20(2):388-400.
13. Moyer VA. Risk assessment, genetic counseling, and genetic testing for BRCA-related cancer in women: U.S. Preventive Services Task Force recommendation statement. *Ann Intern Med*. 2014;160(4):271-281.
14. Screening for osteoporosis: U.S. preventive services task force recommendation statement. *Ann Intern Med*. 2011;154(5):356-364.
15. Helfand M, Carson S. *Screening for Lipid Disorders in Adults: Selective Update of 2001 US Preventive Services Task Force Review*. Rockville (MD)2008.
16. Chen ES, Carter EW, Winden TJ, Sarkar IN, Wang Y, Melton GB. Multi-source development of an integrated model for family health history. *J Am Med Inform Assoc*. 2014.
17. Rosenbloom ST, Denny JC, Xu H, Lorenzi N, Stead WW, Johnson KB. Data from clinical notes: a perspective on the tension between structure and flexible documentation. *Journal of the American Medical Informatics Association : JAMIA*. 2011;18(2):181-186.
18. Johnson SB, Bakken S, Dine D, et al. An electronic health record based on structured narrative. *J Am Med Inform Assoc*. 2008;15(1):54-64.
19. Wilson BJ, Carroll JC, Allanson J, et al. Family history tools in primary care: does one size fit all? *Public Health Genomics*. 2012;15(3-4):181-188.
20. Scheuner MT, de Vries H, Kim B, Meili RC, Olmstead SH, Teleki S. Are electronic health records ready for genomic medicine? *Genet Med*. 2009;11(7):510-517.
21. Green RF. Summary of workgroup meeting on use of family history information in pediatric primary care and public health. *Pediatrics*. 2007;120 Suppl 2:100.
22. Rich EC, Burke W, Heaton CJ, et al. Reconsidering the family history in primary care. *J Gen Intern Med*. 2004;19(3):273-280.
23. Sussner KM, Jandorf L, Valdimarsdottir HB. Educational needs about cancer family history and genetic counseling for cancer risk among frontline healthcare clinicians in New York City. *Genet Med*. 2011;13(9):785-793.
24. Bernhardt BA, Pyeritz RE. The economics of clinical genetics services. III. Cognitive genetics services are not self-supporting. *Am J Hum Genet*. 1989;44(2):288-293.
25. Qureshi N, Wilson B, Santaguida P, et al. *Collection and use of cancer family history in primary care*. US Department of Health and Human Services, Public Health Service, Agency for Healthcare Research and Quality; 2007.
26. Ginsburg GS, Willard HF. Genomic and personalized medicine: foundations and applications. *Transl Res*. 2009;154(6):277-287.
27. Scheuner MT, Sieverding P, Shekelle PG. Delivery of genomic medicine for common chronic adult diseases: a systematic review. *JAMA*. 2008;299(11):1320-1334.
28. Peace J, Valdez RS, Lutz KF. Data-based considerations for electronic family health history applications. *Comput Inform Nurs*. 2012;30(1):37-45.
29. 2 MUS. Meaningful Use Stage 2 - Family Health History. <http://www.healthit.gov/providers-professionals/achieve-meaningful-use/menu-measures-2/family-health-history>.

30. Surgeon General My Family Health Portrait. <https://familyhistory.hhs.gov>.
31. Owens KM, Marvin ML, Gelehrter TD, Ruffin MTt, Uhlmann WR. Clinical use of the Surgeon General's "My Family Health Portrait" (MFHP) tool: opinions of future health care providers. *J Genet Couns*. 2011;20(5):510-525.
32. HL7 Version 3 Implementation Guide: Family History/Pedigree Interoperability, Release 1 - US Realm. [http://www.hl7.org/implement/standards/product\\_brief.cfm?product\\_id=301](http://www.hl7.org/implement/standards/product_brief.cfm?product_id=301).